

October 2020

Got Metadata in Your Future? Lessons Learned from Describing a Unique Image Collection

Scott M. Dutkiewicz
Clemson University Libraries

Jessica Serrao
Clemson University Libraries

Charlotte Grubbs
Clemson University Libraries

Follow this and additional works at: https://scholarcommons.sc.edu/scl_journal



Part of the [Library and Information Science Commons](#)

Recommended Citation

Dutkiewicz, Scott M.; Serrao, Jessica; and Grubbs, Charlotte (2020) "Got Metadata in Your Future? Lessons Learned from Describing a Unique Image Collection," *South Carolina Libraries*: Vol. 4 : Iss. 2 , Article 9.

DOI: <https://doi.org/10.51221/sc.scl.2020.4.2.10>

Available at: https://scholarcommons.sc.edu/scl_journal/vol4/iss2/9

This Conference Proceeding is brought to you by Scholar Commons. It has been accepted for inclusion in South Carolina Libraries by an authorized editor of Scholar Commons. For more information, please contact dillarda@mailbox.sc.edu.

Got Metadata in Your Future? Lessons Learned from Describing a Unique Image Collection

Abstract

This practical session covers how Clemson University Libraries' metadata team describes their largest digital collection of historical images. It focuses on what the team has learned from the project, including developing workflows and strategies for describing images, creating a local heading controlled vocabulary, and leveraging expertise to streamline metadata creation. The team explains the metadata management tool CollectiveAccess, shares examples from the collection, and discusses benefits of documentation. The session concludes with continued metadata challenges.

Keywords

metadata, digital collections, photographs, libraries, archives, research

Got Metadata in Your Future? Lessons Learned from Describing a Unique Image Collection

Figuring out how to get started with your first metadata project can be a daunting task. The intention of this presentation is to cover some of the situations you might face, drawn on three years' experience with a very large photograph collection. The presenters introduce the metadata team structure and the images collection, provide a brief description of Clemson's Digital Collections and CollectiveAccess, describe some strategies for formulating descriptions, explain their local heading controlled vocabulary, touch on the necessity of expertise and documentation, and conclude with current challenges and hopes for future development.

What Is Metadata and Why Is It Important?

There are countless definitions of metadata. The International Federation of Library Associations (IFLA) defines metadata as "any data used to aid in the identification, description and location of networked electronic resources" (International Federation of Library Associations, 2005). There are also many types of metadata. A familiar division is the trio of descriptive, administrative, and structural metadata. All of them are essential to a digital project, but for the purposes of this presentation, and because it is our unique contribution to the entire process, the authors will focus on descriptive metadata.

Because metadata is essential to our professional lives, librarians and archivists instinctively appreciate its value. But if you are new to metadata creation, you might want to hone an elevator talk to explain it to others. One approach is to ask if one shops or banks online or uses email. If so, metadata is involved! Metadata provides vital support to information seeking, especially in our current situation during COVID-19, since many resources are only online. Metadata not only enables the work of researchers and contributes to the preservation of institutional memory.

Background

The Metadata Team

In order to create metadata, one must consider your team and collaborative relationships. The Metadata and Monographic Resources Team is one of five teams in Clemson Libraries' technical services unit. The team comprises two librarians and two metadata specialists. Team members used to catalog only print monographs and special formats materials. Since 2016, work has been focused on descriptive metadata for digital collections.¹ The team does not manage all aspects of Clemson's Digital Collections metadata work. Rather, this is a cross-unit collaborative effort. Special Collections and Archives manage and describe the original source material, and the Library Technology unit's Digitization Lab digitizes the materials and publishes the images and metadata on the WordPress website.

Clemson University Historical Images Collection

Many digitization projects start with small collections. If your organization is just beginning with metadata, that is the best approach. However, the metadata team did not get that luxury. The Clemson University Historical Images collection, which spans more than a century and includes many topics of interest, has some 2,800 images already accessible online, another 3,400 images awaiting description, and thousands more to be digitized.

Clemson's Metadata Environment

To accomplish our metadata work, we utilize several platforms to create, manage, and publish our digital collections metadata and images. We will review the two primary platforms-- the digital collections website and the backend metadata management system.

¹ See Dutkiewicz & Serrao for a case study of the evolution of the cataloging team into a metadata team.

Clemson Digital Collections

Clemson's Digital Collections are displayed using a WordPress site. Take a virtual tour of the site by following along here.² This will give you an idea of what our metadata looks like to users and how metadata affects the browsability of the historical images collection. Navigate to the site by going to <https://digitalcollections.clemson.edu>. On the homepage, if you scroll down, you can see all the collections listed by their content type. Once you click into the Historical Images collection, you come to a landing page with an introductory paragraph describing the collection. Scrolling down, you will see certain topics we chose to highlight, and further down is a geographic map of collection items. This data is pulled from the geographic location field and uses GeoNames terms.

If we want to view everything in this collection, we can scroll back up to the search bar where you see the icon See All. In the list of all items, it is easiest to either search by a keyword or use the facets on the left to narrow down the results. You can narrow by several facets. Collection will not help us here since we are viewing items from only one collection. Creator is the entity responsible for creating the physical photograph and, in this collection, is often a photographer or department on campus. The creator facet pulls from the Related Entities field. You can also narrow by a single date or date range. The topic facet pulls from all our subject heading fields, whether that is Library of Congress Subject Headings and Name Authority Files or our locally created subject headings. Location, just like with the map view, pulls from the geographic location field or the local subject field if it is a building on campus. The Type facet is defaulted to still image since this is a collection of photographs. Medium values are from the Getty Art & Architecture Thesaurus. As you can see, there are redundancies and inconsistencies in the metadata in this field and it needs some work to normalize the values, which is a project we have started. The Rights facet currently use one rights statement: Clemson University Works.

² You can also follow along with the virtual tour screencast video at <https://drive.google.com/file/d/14wID0vPumhvglcm-WIAvggSgevtgMBRY/view?usp=sharing>.

In the future, we would like to adopt more comprehensive and standardized statements like those on RightsStatements.org.

Let us look at an item and its metadata record. We will select Nursing students from the topic facet and see what that pulls up. Clicking into the item, you first see the image viewer where you can zoom and move the image around, as well as download a .jpg. Below the image is the metadata. This is a pretty standard record with a Title, Collections, Dates, Descriptive Information, Locations, Topics, Physical Descriptions, Contributing Institutions, Identifiers, and Rights.

Let us go back and look for another topic. We can deselect our previous selection and click reset to clear it out. Then, let us go to African American students in the topic list. There is only one result. This example illustrates that metadata decisions are constantly evolving. We no doubt have more than one photo of African American students in our collection, but it was not consistently applied. This calls for a larger discussion of how we should equitably describe Black students and students of color and what terms we should use. This is a collaborative discussion we need to have with multiple stakeholders, including experts across the Libraries, our users, and most importantly, people of color in our community. So that is the shiny WordPress front end. Now what does this look like on the backend?

CollectiveAccess

We use CollectiveAccess to manage and preserve our metadata records. It is an open-source web-based collections management program used by many cultural heritage institutions. It is flexible and easy to customize. Because it is web-based, multiple users can be logged in and editing records simultaneously from any location with Internet access. Let us take a walk-through of this platform to show how we create metadata for our digital collections.³ Once logged in to CollectiveAccess, the landing page is a customizable dashboard. I will then select a set of records to look at from the

³ You can follow along with the virtual tour screencast video at <https://drive.google.com/file/d/19WvZriyVTMkW0aWO8RAfWUUZwjP4t4is/view?usp=sharing>.

dropdown list. We manage our projects in chunks by archival container, which means for every box of physical materials we digitize, we create that as a set in CollectiveAccess. Once a set is selected, a results page lists all the items in that box. This list is a grid view of the metadata, which is not the easiest to view because it extends to the right, but it does give a quick glance at what is going on across these items.

If we want to describe the photo with an identifier of ua100_002932, we can navigate to that corresponding metadata record. On the left-hand side next to each record is a page icon that will open the metadata for that single item. Once opened, the object record (aka metadata record) displays all the metadata elements available. There is the title, alternative title, the object identifier that helps us relate our metadata record to the image file, and compound object for objects with multiple pages. The cataloging status lets us internally track the status of completion for each record and who completed it, and the cataloger note is also internal, so we have a place to drop observations and reference information that does not need to be published. The related entities field is that creator facet you saw on the website, and then contributing institutions identify who holds the item, who hosts the item online, and who digitized the item.

The next section is our subject headings. CollectiveAccess pulls vocabularies from LCSH and LCNAF. You can type in a term and get suggestions from which you can browse and select. The local subjects field, covered in more detail later, is a controlled vocabulary we manage and use for locally specific terms. This field acts the same way as LCSH and suggests terms from a controlled list. We then have the medium field which populates from Getty AAT, and geographic location which searches across GeoNames. There is a description field, and fields relating to the format of the digital object, the language of the item, and finally Special Collections supplied metadata and DigiLab notes, which work like the Cataloger Note as internal notes that help us communicate across units.

There are multiple screen views on the left-hand column, and we are currently in the Cataloging Team view, which contains all the fields the metadata team uses. We also customized an SCDL required metadata screen, which contains only those fields the South Carolina Digital Library requires us to provide for aggregation. There is also a Digital Lab screen with descriptive and technical metadata fields that are most used during digitization. The rest of the various screen views are by metadata category that provide quick views by subjects, locations, rights, or URLs. Then CollectiveAccess provides a Summary view of all metadata in the object record and stores a change log of actions taken on the record.

This is a very quick overview of CollectiveAccess, but it gives you an idea of how the system works for Clemson Digital Collections. Most importantly, it provides a user-friendly interface so many hands across multiple units in the Libraries can create metadata.

Strategies for Describing Images

If you have digital projects coming down the line, then you inevitably have metadata in your future. We would like to share some tips we have found useful for describing Clemson's photograph collection. While the following examples will highlight situations specific to Clemson, we hope you gain some insight for your own projects. The biggest takeaway is that there is no "right" metadata, and it is an evolving and adaptive process. We try our best to make descriptive decisions that cater to our researchers' needs, the needs of each collection, and available institutional resources (including time!). Next we will walk through the questions we ask when describing photographs of buildings on campus.

Describing Buildings

One of the most important early sequences in the Historical Images collection were photographs of college buildings. To tackle this, the team lead began defining a strategy of asking questions to help gather the facts needed to provide metadata. It is a matter of learning historical facts and then conveying those facts with metadata vocabularies. First, knowing the history of the building is

important, which is sometimes complex or unknown. Many buildings on the campus changed names, sometimes coincident with a change in the use of the building. A decision was made to use the name of the building that applied *at the time* of the image. For this reason, dates must be carefully noted to decipher the name used. Maps and other resources often reveal clues.

Important historical questions:

- Do you recognize and know the name of the building?
- When was the photo taken?
- Has the name changed over time?
- Where is the building on a map?
- Do you know the type and use of the building?
- Has use changed over time?
- Once the facts are known (as many as can be ascertained) the next step is to communicate those facts in the metadata record. Do vocabularies include the building?

Very few Clemson University buildings are in the Library of Congress Name Authority File. For instance, the main administrative building dating to the founding of the college is not in LCNAF! This situation is likely to be the same for your institution. Early in the project, the team lead compiled a local list of campus buildings with a preferred access point. This list eventually joined the list of local headings in CollectiveAccess (more on local headings below).

Important vocabulary questions:

- Is the building in the Library of Congress Name Authority File?
- Are you using a generic heading across the collection? (LCSH: College buildings)
- Is the building listed in GeoNames?

For general collections, this information will probably suffice, but institutions with specialized collections might want to enhance the record with additional access points. One can also provide

metadata for entities, such as architect, builder, and donor; dates for the entire building lifecycle such as groundbreaking, construction, dedication, opening, renovation, and demolition. One could note the architectural style, and, if a dwelling, the current or former owners and occupants. Just as one would not include a complete biography in a record for an image of a person, resist the temptation to include a full building history into the Description field of every image for a building. If known, this information should be input into the authority record(s) for the building.

With every image one studies, often the first experience is disorientation. You are placed back in time into a world in which you are a stranger. This experience may be unsettling at first, but ultimately, very satisfying because there is always something new to learn. That is why some of us get hooked on metadata.



Aerial view of Clemson House, 1978, Clemson University Photographs, Series 100, Special Collections, Clemson University Libraries, Clemson, SC

Metadata is an attempt to provide ready-made solutions to common information needs. But situations occur that require us to revisit and revise metadata. One would expect that metadata for buildings is rather static. However, on Sunday, December 3, 2018, Clemson House, originally built as

a motel and had been adapted into a dormitory, was imploded to clear the location for another campus building. The next Monday morning, the team lead arrived at the office and said, “We have to do something in the metadata to show that Clemson House is no longer standing.” The metadata team

strategized and decided to apply an LC subject term of Lost architecture. That allows the searcher to find any images of structures no longer extant.

Buildings and campus features are fascinating elements in Clemson's history, but even more important are the people of Clemson who envisioned, built, lived in, and taught in those buildings. The next section will cover the description of people in photographs.

Describing People

We realized very quickly that describing people in photographs is not as straightforward as it may seem. Many questions needed to be answered. How do we balance the information from the original photograph with a normalized and consistent description that can ease the user's experience of the materials? There are so many ways to identify which person is which in each photo, and because there are multiple people creating metadata for this project, we wanted to ensure consistency across the collection.

To do this, we wrote guidelines for formatting Titles and Descriptions of people in photos.

- Keep formatting consistent across collection(s).
- Focus on describing the known and not the unknown.
- Do not use question marks (?) as placeholders for unidentified persons. Instead use "unidentified."
- For large groups with many known names, follow below formatting:
 - Left to right - Peter Jones, Fred Lee, Olive Lee, Joe Johnson, President James Barker
 - Back row left to right - John Smith, Jane Doe, A. B. Smith, T. L. Jones
- For large groups with few known names, follow parenthetical formatting:
 - President James Barker (third from right) with students
 - Peter Jones (seated center)

- For small groups with most or all people identified, follow parenthetical formatting or list with directionals:

- John Smith (left) with Jane Doe
- Fred Lee (standing left), Olive Lee (seated), and Joe Johnson (standing right)
- On back: Left to right - A. B. Smith, T. L. Jones, John Hughes

We have formatted how to list names out and how to identify individuals using parentheses with directionals. If it is only one or two people, this description might become the title. If it is a group photo, this information will go into the description field and we would apply a more generic title, such as “Cadets posing on steps of Agricultural Hall.”

Let’s walk through an example. This is a group photograph of cadets from 1913 showing the final metadata formatted according to our guidelines. First, we rely on any information the photo provides.

Title	Group portrait of the Big Five cadets at the fair, 1913
Description	Left to right - Charles John "Tin" Shannon, Walter A. "Walt" Bigby, Monroe Boyce "Judge" Bailey, Tom "Major" Spratt, and Warren Thomas "Cap" White. On back: The "Big Five" at the "Fair"
Local Subjects	Shannon, Charles John; Bigby, Walter A. (Walter Ashbury), 1894-1973; Bailey, Monroe Boyce, 1894-1976; Spratt, Tom; White, Warren Thomas, 1895-1973

**LC Subject
Heading**

Military cadets



Group portrait of the Big Five cadets at the fair, 1913, Clemson University Photographs, Series 100, Special Collections, Clemson University Libraries, Clemson, SC.

In this case, we received a transcription of what was written on the back of the photograph from Special Collections and Archives. It read: From left to right; "Tim" Shannon; "Walt" Begley [illegible]; "Judge" Bailey; "Major" Sjratt [illegible]; "Lafa" White [illegible]; The "Big Five" at the "Fair" 1913. You can see it contains a list of names and nicknames in quotations with several sections of illegible text. After the names, there is what appears to be an original title applied by the owner or creator - "The Big Five at the Fair 1913." We used this to formulate the title, adding in "group portrait" and "cadets" for further context. We also applied the LC Subject Heading of Military cadets, which is the term we chose to assign to all photographs depicting Clemson cadets. Easy, right?

Next was the challenge of the names. To avoid creating a mile-long title, we put the names instead in the Description field. We needed to do some digging into university archives to decipher the

illegible portions. Many of the names were found in Clemson's yearbooks,⁴ which thankfully lists full names and often nicknames of seniors. We even found a reference to their membership in the "Big Five" so we knew we had the right people. We provided the fuller names in the description field and retained their nicknames. To aid researchers who may wish to find more materials on these students, and because they were not represented in LCNAF, we created local subject headings as access points for each. You can see, some of the headings are more complete than others with birth and death dates. We could research all day into each of these names, but for the sake of time, we decided to limit our searches unless we need to disambiguate between two similar names and need further details to determine who is who. If we quickly find birth and death dates, we will include them, but they aren't always found easily in the limited search time we have. All these local headings for the individuals then become linked topics on the WordPress site. Researchers can find all materials for an individual with a single click.

Local Headings

Local subject headings are standardized descriptive access points we assign to digital objects that are unique to Clemson University's collections. A local subject heading is assigned if no suitable heading to describe an object is found in a nationally authorized list, such as the Library of Congress Name Authority File (LCNAF) or Library of Congress Subject Headings (LCSH). Local subject headings are created to improve discovery of our digital materials and provide access points that cover people, corporate bodies (which include structures), and events represented in our digital collections.

Reasons for creating a local subject heading include:

- If the subject appears either frequently within one collection or across all digital collections. In that case, creating a local subject heading can help you avoid having to describe the same

⁴ Clemson yearbooks are digitized and available online, which facilitates text searching to find student names. See https://tigerprints.clemson.edu/ua_pubs_yearbooks/ to view this collection.

subject repeatedly within the item record itself.

- If the subject is a person, corporate body, or event that is unlikely to be featured in another institution's collection.
- If you are considering implementing linked open data in the future.

We strongly recommend establishing rules for the creation of local subject headings at the beginning of your metadata project. Deciding what type of subjects demand the creation of a local heading and then consistently applying that standard will save you time down the road.

Managing Local Headings

Successfully utilizing local subject headings requires managing a controlled vocabulary, which is "an organized list of words and phrases used to index content and retrieve content through browsing and searching." (Harpring, 2010, p. 12) The most basic of controlled vocabularies is the term or headings list, a list of unique words or phrases used to identify a specific person, corporate body, event, or other subject. A more complex type of controlled vocabulary are authority files, which "include cross-references for variant or alternate terms. Authority files often include other contextual or biographical information to assist users with disambiguation." ("Metadata and Discovery @ Pitt," 2018) Our own homebrewed controlled vocabulary is a little bit of both - while we do not use alternate terms for local subject headings, we do often include additional context or information about a subject in the local subject heading record.

Managing a controlled vocabulary of local subject headings, especially a term list, can be as simple as maintaining a spreadsheet with every heading - provided that those headings are applied consistently across the collection or collections. When we create a local subject heading in CollectiveAccess, we can retrieve that heading while working in a record just by searching for the correct term in the Local Subject Heading field. CollectiveAccess also enables us to manage our controlled vocabulary, as shown in the screenshot below. By going to Lists & vocabularies and selecting Local

Subjects, we can pull up the complete list of local subject headings used across the Clemson Digital Collections cataloged by our metadata team, including Historical Images, the Strom Thurmond papers, the A. Wolfe Davidson collection, and the James Byrnes collection. Being able to view the entire list of local subject headings is important because it helps us weed out duplicate and mis-formatted terms - something we previously struggled with.

The screenshot shows a web form for creating subject headings. It is divided into two main sections: 'Preferred labels' and 'Identifier'. In the 'Preferred labels' section, the 'Item name (singular)' field contains 'Wylie, William Harry, 1888-1931' and is highlighted with a red box, with the text 'Subject name' in red to its right. Below it, the 'Item name (plural)' field also contains 'Wylie, William Harry, 1888-1931'. The 'Description' field contains a detailed text entry about the subject's identification and source, highlighted with a green box, with the text 'Subject description, including original image and additional identification sources.' in green below it. The 'Identifier' section at the bottom has a 'Unique identifier' field containing 'sccleucl_lh_0000914', which is highlighted with a blue box. A warning icon and text are visible next to this field.

Preferred labels

Item name (singular)
Wylie, William Harry, 1888-1931 Subject name

Item name (plural)
Wylie, William Harry, 1888-1931

Description
First identified ua100_003247. Full name and place of birth taken from Clemson College Annual 1907. Birth and death dates taken from Find a Grave, via Ancestry: <https://www.findagrave.com/memorial/169122268>.
Subject description, including original image and additional identification sources.

Identifier Unique identifier
sccleucl_lh_0000914 ⚠ Changing this value may break parts of the system configuration

This screenshot provides an overview of every piece of information that goes into the creation of our local headings. Subject name--people, corporate bodies, and events--formatted using RDA standards, followed by subject description, and a unique identifier formatted using the MARC organization code for Clemson University, R. M. Cooper Library. This identifier is consistent across all our digital projects in CollectiveAccess, not just the Historical Images collection.

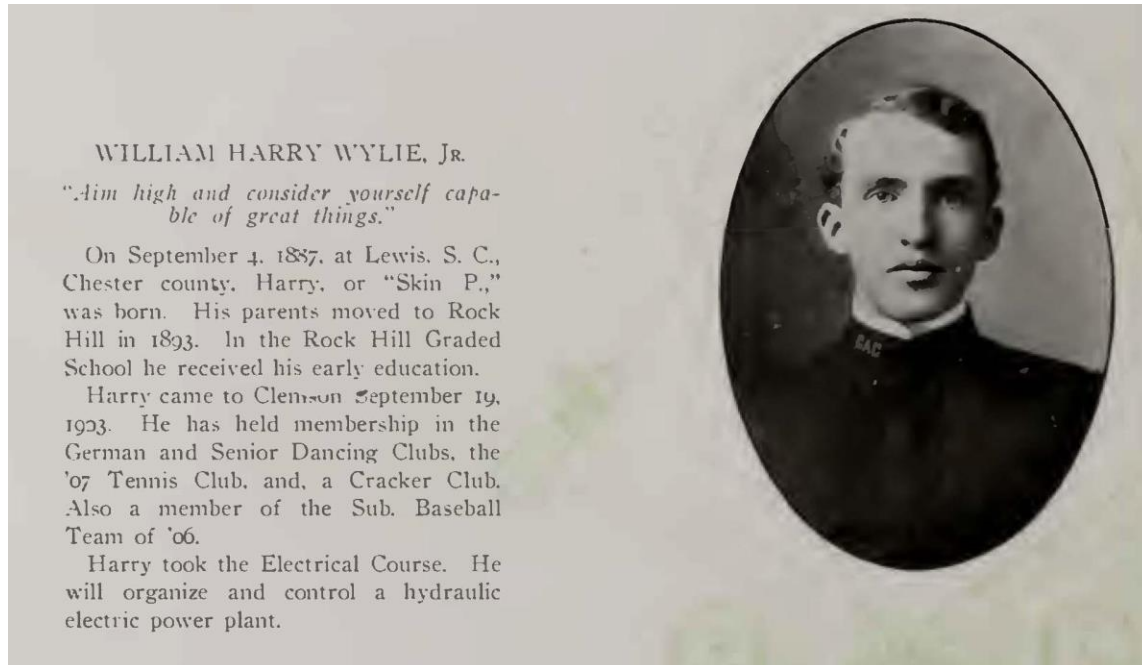
The description field identifies the first image assigned this local heading, as well as information about the subject which would not be appropriate to include on an item record level but is still crucial for the metadata practitioner to know in order to consistently apply the heading. This might include

extraneous biographical details of a person, historical context of building name usage, or organizational sponsorship of an event. The description field may also include source citations if we used additional resources to identify a subject. What information to include here is up to the metadata practitioners' discretion, but it may be useful to keep an eye toward using this data to create an LCNAF or sharing it with other libraries and cultural institutions in the future.

Subject Identification Resources

At times, as a metadata practitioner you may not have enough information about an image or its subject in order to create and properly format a local subject heading. For instance, you may have only a partial name of a person, or not know their birth or death dates. When it comes to identifying subjects and creating local headings, it is important to have reliable, trustworthy sources of information. Some of these resources will be unique to your library, including institutional records, publications, and databases. For our Historical Image collection, Clemson-specific resources have been invaluable in helping identify subjects for local headings. These resources include Clemson's yearbook publications; the student newspaper, *The Tiger*; Board of Trustees meeting minutes and annual reports; historical and current maps; and other university publications. All these sources have been digitized and are available on our institutional repository, TigerPrints.

Let's walk through an example. For one local heading, we only had a partial name, "W. H. Wylie," as transcribed from handwritten notation on the back of a group photograph. Using the resources available on TigerPrint, we were able to locate a William Harry Wylie, Jr. in the Clemson College Annual (a precursor of the Taps yearbook) whose dates of attendance and graduation from



Clemson University, "Clemson College Annual (1907)." Yearbooks. 8. <https://tigerprints.clemson.edu/yearbooks/8>.

Clemson matched the year the photograph was taken.

Thanks to this resource, we have a full name and an approximate place and date of birth.

(Approximate because our experience with student publications has been that dates are sometimes inconsistent, as demonstrated in this example. Again, choose which resources you rely on for metadata with care.) This is enough information to perform a search using Ancestry Institution. Ancestry is a genealogical website that pulls records from both proprietary and government databases, and Clemson Libraries subscribes to the institution version. A quick search on Ancestry turned up multiple records for William Harry Wylie, Jr. of Rock Hill, South Carolina, including a listing from Find a Grave with birth and death dates. While the September 4th, 1888 birth date is not an exact match to Mr. Wylie's biography in the Annual, based on the information available, and considering the reliability of the resources, we feel comfortable that 1888 is the correct year. From there, we have all the information needed for a complete local subject heading.

One thing to keep in mind when deciding on best practices for your metadata project is how much time you want your metadata practitioners to spend researching local headings. As you can

imagine, trawling through archival material to find the full names and birth and death dates for every individual in a group photograph of a dozen or more is very time consuming. Here at Clemson, we err on the side of completion, but if full names and dates are not forthcoming after checking with our go-to identification sources and performing a basic database search, we move on. What searches are worth pursuing and what are just rabbit holes of bad data is something you will develop a feel for after some experience working in metadata.

Leveraging Expertise and Documentation

One way to save time and resources is to rely on experts with subject and collection knowledge to help create metadata and reduce redundant work. In the case of the Historical Images collection, the metadata team works closely with Special Collections and Archives both to take advantage of our archivists' extensive knowledge of Clemson history, and to better streamline our descriptive workflows. One example of this relates to local subject headings. We decided to create headings for cadet military units since this is a prominent part of university history and researcher interest. This is a subject the metadata team knew little about, but which Special Collections and Archives were experienced in. They could more quickly create the list of military units and save the metadata team's time. Similarly, we shifted the metadata workflow to rely more on Special Collections and Archives' initial item descriptions created during processing. Now, metadata records reach the metadata team in a more complete state with titles and barebones descriptions already in place. This enables the metadata practitioners to focus on adding access points and creating local subject headings as necessary.

Because three different library units are involved with our metadata projects (Special Collections and Archives, Technical Services, and Library Technology), documentation is essential, so everyone understands the rules. Unlike most traditional cataloging in which the rules are, for the most part, predefined--like the Anglo-American Cataloging Rules (AACR), Resource Description and Access (RDA), and Library of Congress Subject Heading application--metadata documentation is an ever-evolving

process. The key to consistency is to write down and follow your standards. The two most important documents that Clemson's Digital Collections use are the Project Charter and the Application Profile.⁵ Together these documents form a foundation for action, like the bylaws of an organization. Like bylaws, they are sufficiently stable to guide routine decisions, but must be flexible to allow for amendment as important issues come up.

Continued Challenges and Future Directions

Like any metadata story, there are still challenges we struggle with. Selecting appropriate subject headings is an ongoing discussion. As we work through this massive collection of photographs, we come across new subjects and continue to have conversations over which terms will be more inclusive and help researchers pull like materials together. This relates back to our African American student example with one search result. We need to have these tough conversations of what terms to use to describe students and employees of color, then apply them consistently to optimize searching.

Entities are another challenge. We currently use the Related Entities field in CollectiveAccess to assign creators. This is in the same shape our local headings were in a year ago. It is uncontrolled with duplications, inconsistent formatting of names, and general messiness. This will be a future clean-up project, which will also entail creating guidelines to improve consistency. We also need to ask the question - what do we use entities for and how do they differ from our local headings? Are they just for assigning creators while local headings are for subjects of a photo? Entities represent people and corporate bodies that may also be represented in our local headings or in Special Collections and Archives' ArchiveSpace instance. How do we get these various lists to talk to one another to reduce redundancies and eliminate duplicate work?

⁵ See Clemson Libraries Digital Project Charter template, <http://bit.ly/CUL-Project-Charter>, and Clemson Libraries Metadata Application Profile, <http://bit.ly/CUL-Metadata-Application-Profile>.

We also continue to face the challenge of photos with little to no context. These usually have nothing written on the back, and titles need to be derived from what is in the photograph. In our new workflow, Special Collections will help generate titles using their historical expertise, which will reduce this challenge for the metadata team. But we still need to apply appropriate subject headings without making assumptions of what the photo is about. For example, we cannot assign a geographic location of Clemson University if there is no identifying information in the photograph of where it was taken. Just because it is a Clemson University historical image does not mean that is where the photograph is geolocated.

One last major challenge is our metadata revision process. If revisions need to be made to existing records, it is currently a manual process that involves multiple humans. Revisions are made in CollectiveAccess, then documented in a Google spreadsheet. From there, each metadata record is exported from CollectiveAccess, indexed, and then updated on WordPress. Our aspiration is to create a more automated and responsive process where our currently siloed platforms speak to one another and there is less reliance on multiple human inputs.

And finally, future directions. We would like to conduct more usability testing to see how researchers navigate our digital collections. This will help us create informed metadata approaches based on user data and not based solely on our own assumptions or inclinations. We would also like to pursue linked data projects to broaden our collections' discoverability and research capabilities. This could enhance the way researchers see relationships between our materials, between our collections, and across other institutions' collections. As a first step, we began assigning linked data URIs for LC authorities in CollectiveAccess and rely on vocabularies that support linked data. We would also like to explore the capabilities of crowdsourcing to enhance our metadata with community expertise and community-driven descriptions.

Conclusion

We hope this presentation has given you ideas for how to tackle your own metadata projects and directions for what the future of metadata looks like. If anything, you can take away two major points from this presentation: There is no “wrong” way to metadata, as long as you strive for consistency and interoperability. Every institution must find what works for them. And metadata is never done - there is always clean-up, enhancements, and revisions. If we strive to support new and emerging ways our researchers are searching for and using our materials, we can become part of the necessary metadata evolution and improve access to our digital collections.

References

- Aerial view of Clemson House, 1978, Clemson University Photographs, Series 100, Special Collections, Clemson University Libraries, Clemson, SC, <https://digitalcollections.clemson.edu/single-item-view/?oid=CUIR:74BDA2D48AC043B63781D54C6239F4EA>.
- Clemson Digital Collections. <https://digitalcollections.clemson.edu/>.
- Clemson Libraries Digital Project Charter template. <http://bit.ly/CUL-Project-Charter>.
- Clemson Libraries Metadata Application Profile. <http://bit.ly/CUL-Metadata-Application-Profile>.
- Clemson Libraries. (n.d.). *Yearbooks*. TigerPrints. https://tigerprints.clemson.edu/ua_pubs_yearbooks/.
- CollectiveAccess. <https://www.collectiveaccess.org/>.
- Dutkiewicz, S. M. and Serrao, J. L. A road taken: A cataloging team becomes a metadata team. In S. Marien (Ed.), *Library technical services: Adapting to a changing environment* (pp. 65-81). Purdue University Press, 2020.
- Find A Grave. (2020). <https://www.findagrave.com/>.
- GeoNames, <http://www.geonames.org/>.
- Getty Research Institute. (2017, March 7). *Art & Architecture Thesaurus Online*. <http://www.getty.edu/research/tools/vocabularies/aat/>.
- Group portrait of the Big Five cadets at the fair, 1913, Clemson University Photographs, Series 100, Special Collections, Clemson University Libraries, Clemson, SC.
- Harpring, Patricia (2010). *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and other Cultural Works*. Getty Publications.
- International Federation of Library Associations. (2005, October 24). Digital Libraries: Metadata Resources. <https://www.ifla.org/node/9337>.
- Rights Statements. (n.d.). <https://rightsstatements.org/>.

Serrao, J. L. (2020, May 27). *Clemson CollectiveAccess tour* [Screencast video].

<https://drive.google.com/file/d/19WvZriyVTMkW0aWO8RAfWUUZwjP4t4is/view?usp=sharing>.

Serrao, J. L. (2020, March 10). *Clemson digital collections tour* [Screencast video].

<https://drive.google.com/file/d/14wID0vPumhvglcm-WIAvgqSqevtgMBRY/view?usp=sharing>.

University, Clemson, "Clemson College Annual (1907)" (1907). Yearbooks. 8.

<https://tigerprints.clemson.edu/yearbooks/8>.

University of Pittsburgh Library System. (2018, August 31.) *Metadata and Discovery @ Pitt: Taxonomies and Controlled Vocabulary*.

<https://pitt.libguides.com/metadataanddiscovery/controlledvocabularies>